ELSEVIER

Journal of Chromatography A, 806 (1998) 279–291

JOURNAL OF CHROMATOGRAPHY A

Enrichment and purification of proteins of *Haemophilus influenzae* by chromatofocusing

Michael Fountoulakis[a,*], Hanno Langen[a], Chris Gray[b], Béla Takács[b]

[a]*F. Hoffmann–La Roche Ltd., Pharma Division, Preclinical Central Nervous System Research—Gene Technology, Building 93-444, 4070 Basel, Switzerland*

[b]*F. Hoffmann–La Roche Ltd., Pharma Division, Infectious Diseases, 4070 Basel, Switzerland*

Received 23 October 1997; received in revised form 15 January 1998; accepted 19 January 1998

Abstract

Haemophilus influenzae is a bacterium of pharmaceutical interest of which the entire genome has been sequenced. Identification of low-abundance proteins in a two-dimensional map is important for the detection of new drug targets. We applied chromatography on Polybuffer Exchanger (chromatofocusing) in order to fractionate and enrich *H. influenzae* proteins, possibly low-copy-number gene products, from larger volumes. Two proteins, major ferric iron-binding protein (HI0097) and 5′-nucleotidase (HI0206) were obtained in pure form and hypothetical protein HI0052 was purified to near homogeneity by this single purification step. Four other proteins, aspartate ammonia lyase (HI0534), peptidase D (HI0675), elongation factor Ts (HI0914) and 5-methyltetrahydropteroyltriglutamate methyltransferase (HI1702), were strongly enriched so that chromatography on Polybuffer Exchanger can be used as an initial step for their isolation. Approximately 125 proteins were identified in the fractions collected from the column. Seventy of these were for the first time identified after chromatography on Polybuffer Exchanger. The proteins enriched by the chromatofocusing step include both low-abundance as well as high-copy-number gene products. They do not belong to a single protein class and the majority of them are enzymes with various functions. The results include a list and a two-dimensional map of the proteins enriched by chromatofocusing. They may be useful in the search of drug targets and in the design of purification protocols for the isolation of homologous proteins from related microorganisms. © 1998 Elsevier Science B.V.

Keywords: Haemophilus influenzae; Chromatofocusing; Protein

1. Introduction

Haemophilus influenzae is a Gram-negative bacterium of pharmaceutical interest, commonly isolated from the respiratory tract. Recently, the complete genome sequence of the microorganism was determined, comprising approximately 1740 open reading frames [1]. The visualization by two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) and the identification of the gene products expressed may contribute to the detection of new drug targets, an important process for the development of new pharmaceuticals [2–4]. Approximately 600 spots representing proteins from the soluble fraction of the microorganism can be visualized on a Coomassie-stained 2D gel and of these about 290 proteins have been identified thus far, partially after enrichment by affinity chromatography on heparin–Actigel [5–7].

In order to identify additional proteins of *H. influenzae* to possibly complete the 2D map, we

*Corresponding author.

0021-9673/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0021-9673(98)00057-0

followed three approaches: (i) enrichment of the number of proteins in the crude extract by using different detergents and chaotropes (manuscript in preparation), (ii) improvement of the sample application techniques for the preparation of the 2D gels [8] and (iii) enrichment of proteins, possibly low-abundance ones, from larger volumes by chromatographic techniques.

The heparin chromatography that we used previously enriched many low-abundance proteins, however, not exclusively. About 40% of the proteins bound to the heparin matrix were nucleic acid-binding proteins [5]. The remaining proteins bound to the affinity gel probably by ionic interactions as heparin carries sulfate groups and consequently acts as an ion-exchanger.

Here we applied chromatofocusing in order to fractionate and enrich proteins of *H. influenzae*. Chromatofocusing is an ion-exchange chromatography. The proteins are bound to the gel matrix, the Polybuffer Exchanger, and are eluted with a specific buffer, the Polybuffer, in the order of their decreasing isoelectric points. Proper choice of the pH values of the equilibration and elution buffers and of the dimensions of the column can result in an efficient protein concentration and high resolution [9–11].

Three proteins of *H. influenzae* were purified to near homogeneity by this single step and at least four other proteins were strongly enriched so that chromatofocusing can be used as a first step for their isolation. The proteins that bound to the Polybuffer Exchanger were separated by 2D-PAGE and identified.

2. Experimental

2.1. Materials

Polybuffer Exchanger 94, Polybuffers pH 7–4 and 9–6, and immobilized pH gradient strips were purchased from Pharmacia. Reagents for the preparation of the one-dimensional- (1D) and 2D-polyacrylamide gels were from Bio-Rad and Serva.

2.2. Protein extraction

H. influenzae cells, strain Rd KW20, were grown in a 10 l fermenter containing 2% fetal calf serum and 10 mg/l hemin as well as the necessary nutrients and salts [5,7]. The fermentation yielded approximately 70 g of wet biomass which was kept frozen at −70°C until use. Sixteen grams of wet cell paste were mixed with 32 ml of 50 m*M* Tris–HCl, pH 8.0, containing 1 m*M* $MgCl_2$, 100 U/ml aprotinin, 5 m*M* ε-aminocaproic acid and 0.2 m*M* dithiotreitol. Nucleic acids were hydrolyzed by addition of 250 U/ml benzonase (Merck). Cells were disrupted in a French press (SLM Instruments) at $1.33 \cdot 10^8$ Pa. $EDTA \cdot Na_2$ was added to the lysate to 5 m*M* final concentration. The mixture was centrifuged at 3000 *g* for 20 min to sentiment intact cells and cell debris. The supernatant was centrifuged further at 18 000 *g* for 60 min to remove cell envelopes. The supernatant from this last step was recovered and centrifuged at 150 000 *g* for 90 min to sediment cell membranes. From the last centrifugation step, a part of the supernatant containing about 200 mg of soluble cytoplasmic proteins was dialyzed against 25 m*M* Tris–acetate, pH 8.3 to remove salts.

2.3. Chromatofocusing

Following dialysis, the soluble protein fraction was filtered through a 0.22 μm pore-size membrane and applied onto a 187 ml (93×1.6 cm) Polybuffer Exchanger 94 column, equilibrated with 25 m*M* Tris–acetate, pH 8.3 at 1 ml/min. The column was washed with 200 ml of the same buffer and the proteins were eluted with Polybuffer, pH 5.0 (800 ml). The elution buffer was prepared by diluting a mixture of Polybuffer 96–Polybuffer 74 (3:7, v/v) tenfold with water and adjusting the pH to 5.0 with acetic acid. The column was further eluted with a linear gradient of increasing salt concentration from 0 to 2 *M* NaCl in 25 m*M* Tris–acetate, pH 8.3. Fractions of 8 ml were collected and pooled according to the elution profile. Nineteen pools were formed which were concentrated to about 1 ml by centrifugation at 2000 *g* in a Millipore Ultrafree-15 device with Biomax-10 membrane and the concentrates were analyzed by 1D- and 2D-PAGE.

2.4. 2D-PAGE

The 2D-PAGE analysis was performed as de-

scribed [5–8,12]. In short, comparable protein sample amounts were applied on immobilized pH 3–10 nonlinear and 6–11 linear gradient strips and isoelectric focusing was performed at 5000 V for 48 h. For the second dimension, the proteins were separated on 9–16% linear gradient polyacrylamide gels at 40 mA/gel. The gels were stained with colloidal Coomassie blue (Novex) for total protein detection and were destained with water.

2.5. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS)

The MALDI-MS analysis was performed as reported [13]. The protein spots were in-gel digested. The gel pieces were destained with 50% acetonitrile in 0.1 *M* ammonium hydrogen carbonate and dried in a vacuum speed evaporator. The dried gel fragments were reswollen in 10 μl of 3 m*M* Tris–HCl, pH 8.0, containing 1 μg of endoproteinase Lys-C (Wako) and incubated at 37°C for 2–16 h. Five μl of 30% acetonitrile, containing 0.1% trifluoroacetic acid were added. After sonication for 3 min, 1 μl of the peptide extract was applied onto 0.5 μl of air-dried matrix. The matrix was prepared by dissolving 10 mg nitrocellulose (Bio-Rad) and 20 mg α-cyano-4-hydroxycinnamic acid (Sigma) in 1 ml of acetone–isopropanol (1:1, v/v). The samples were analyzed on a time-of-flight mass spectrometer (PerSeptive Biosystems) equipped with a reflectron. An accelerating voltage of 20 kV was used. Calibration was external to the samples. The masses found were matched to the theoretical peptide masses of the proteins of *H. influenzae* using the PeptideSearch software [14] or the MS-Fit software at the UCSF Mass Spectrometry Facility accessed via the WorldWideWeb (http://rafael.ucsf.edu/msfit.htm).

3. Results

3.1. Polybuffer exchanger chromatography

The total soluble fraction of the proteins of *H. influenzae* was used for fractionation on Polybuffer Exchanger. After the sample application, the column was washed with one volume of equilibration buffer. Essentially, no proteins were detected in the flow-through or the wash fractions. The column was eluted with Polybuffer, pH 5.0. The collected fractions were pooled according to the elution profile (Fig. 1A) and the pools were first analyzed by sodium dodecyl sulfate (SDS)-PAGE (Fig. 1B; the numbers correspond to the pools). Some of them (such as pools 4 and 9) contained only small amounts of proteins (Fig. 1B, pools 4 and 9), in addition to colored absorbing substances (pool 9). As seen from the SDS-PAGE analysis, the pools of the fractions eluted with Polybuffer contained only few proteins (Fig. 1B, pools 3–15), in particular, the first eluted pools (pools 3–10). In order to recover additional proteins, the column was further eluted with a salt gradient. This treatment yielded a larger number of proteins (Fig. 1A, peaks 16–19 and Fig. 1B, pools 16–19). The column was finally washed with 0.1 *M* HCl which resulted in the elution of few proteins only (data not shown). The pools of the fractions collected from the Polybuffer Exchanger were further analyzed by 2D-PAGE, a technique capable of separating many hundreds of proteins and enabling their identification by mass spectrometry and amino acid composition analysis [15–17]. The 2D gels of selected pools from the chromatofocusing column are shown in Figs. 2–4.

3.2. Purification of proteins of H. influenzae by chromatofocusing

2D-PAGE analysis of pool 1 revealed the presence of one protein spot (Fig. 2A; 1D gel analysis not shown). The protein was identified by MALDI-MS as major ferric iron-binding protein (HI0097). A second minor spot, close to the major one, represents the same protein (Fig. 2A). The heterogeneity is most likely due to carbamylation caused by the high concentrations of urea used during the 2D gel electrophoresis. No other protein spots were detected, indicating that the major ferric iron-binding protein was recovered in a homogeneous form.

Pool 3 included one major protein migrating at approximately relative molecular mass (M_r) 28 000 and a minor one migrating at M_r 60 000 (Fig. 1B, pool 3). The 2D-PAGE analysis showed the presence of three protein spots. Two of them represent hypothetical protein HI0052 and the third spot represents 5′-nucleotidase (HI0206) (Fig. 2B). Pool 6 contained

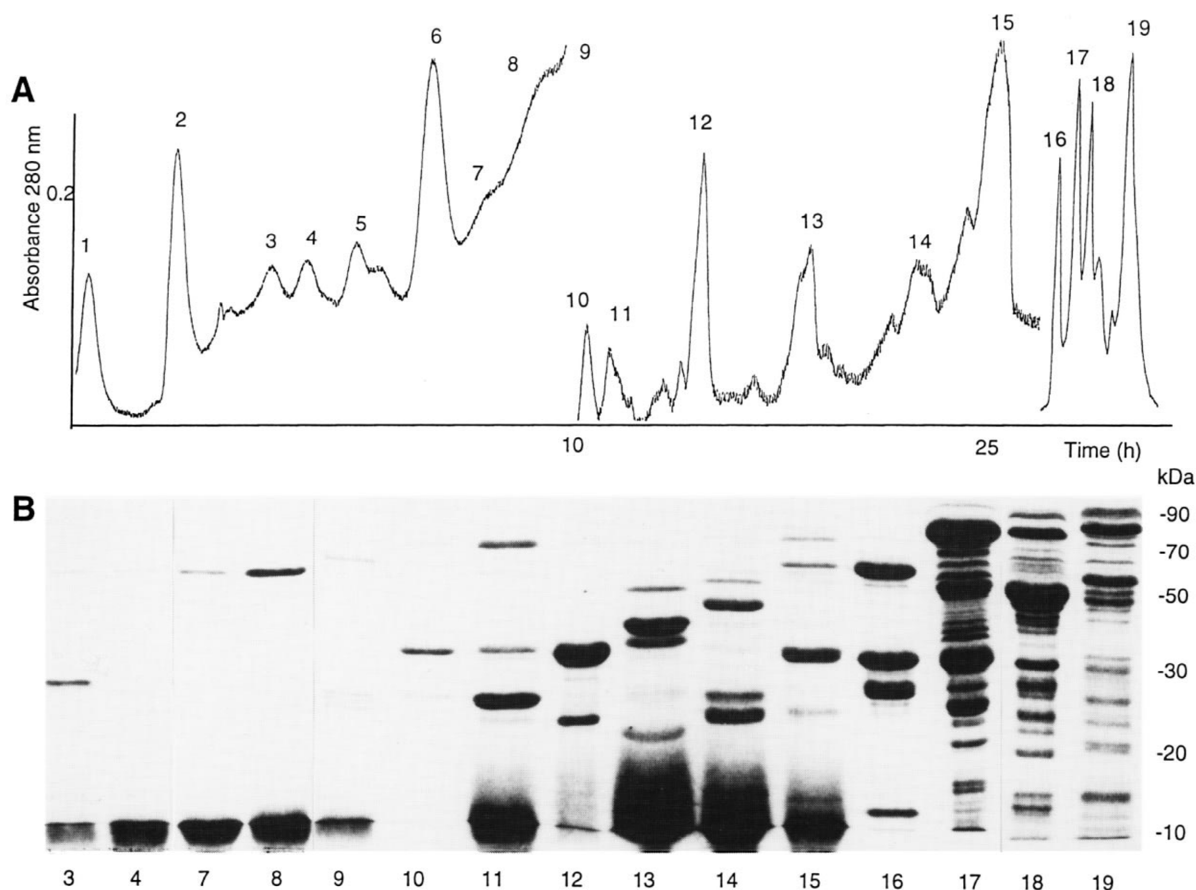

Fig. 1. Partial protein elution profile (A) and SDS-PAGE analysis (B) of the *H. influenzae* proteins eluted from the Polybuffer Exchanger column. The soluble protein fraction was applied onto Polybuffer Exchanger and the column was eluted as stated in Section 2. (A) The parts of the elution profile which include protein peaks are shown. The numbers above the peaks indicate the pools formed by mixing the fractions included in the corresponding peaks. Pools 1–15 were eluted with Polybuffer, pH 5.0 and pools 16–19 were eluted with a NaCl gradient (different chart speeds were used for absorbance recording). (B) The pools were analyzed by 15% SDS gels under reducing conditions. The gels were stained with Coomassie blue. The broad band comigrating with the front (lanes 3–15) corresponds to Polybuffer. The numbers correspond to the numbers of the pools (pools 1, 2, 5 and 6 are not shown).

only 5′-nucleotidase (Fig. 2C; the minor spot at the basic side of the major one represents the same protein). Pool 8 contained larger amounts of 5′-nucleotidase, in comparison with pool 6, but in addition, it contained spots corresponding to D-ribose-binding periplasmic protein (HI0504) and to thiol:disulfide interchange protein (HI1213) (data not shown).

Therefore, in the first pools, collected from the Polybuffer Exchanger three proteins of *H. influenzae* were purified to homogeneity or near homogeneity. Since only the pools were analyzed and not the individual fractions, it is possible that some of the fractions before pooling contained the protein HI0052 in even higher purity. The broad band which comigrates with the front in the 1D gels (Fig. 1B, lanes 3–15) corresponds to Polybuffer which binds Coomassie blue. Polybuffer can be removed from the protein solution by gel filtration chromatography or ammonium sulfate precipitation. Of the three proteins purified by chromatofocusing, the major ferric iron-binding protein was not significantly enriched in comparison with the starting material, whereas the 5′-nucleotidase and the hypothetical protein HI0052

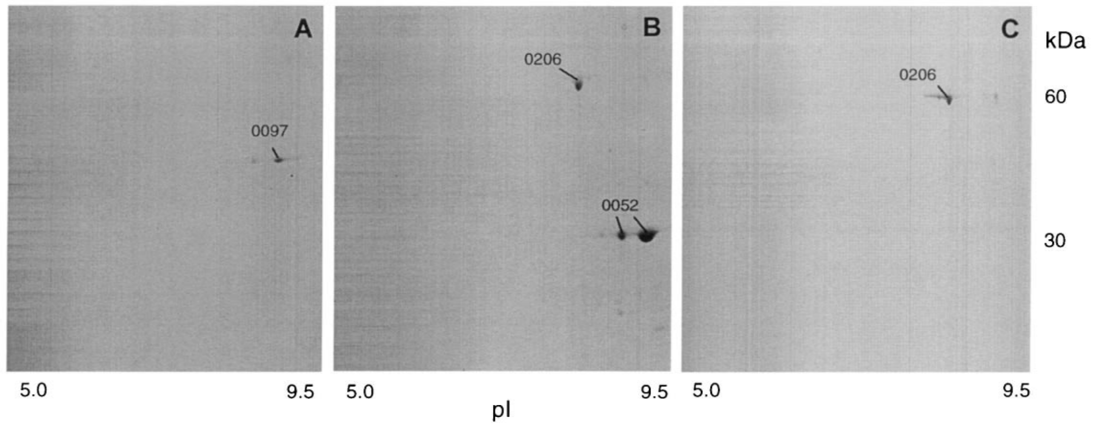

Fig. 2. 2D-PAGE analysis of pools 1 (A), 3 (B), and 6 (C) of fractions collected from the Polybuffer Exchanger column. The proteins were eluted with Polybuffer, pH 5.0 as described in Section 2. The samples were analyzed on 3–10 nonlinear immobilized pH gradient strips, followed by 9–16% gradient SDS gels. The gels were stained with colloidal Coomassie blue and destained with water. The destained gels were scanned in a Molecular Dynamics Personal Densitometer and the images were processed using Adobe Photoshop and PowerPoint software. The proteins were identified by MALDI-MS. The numbers next to the protein spots indicate the HI (*H. influenzae*) identification numbers [1]. The names of the corresponding proteins are given in Table 1.

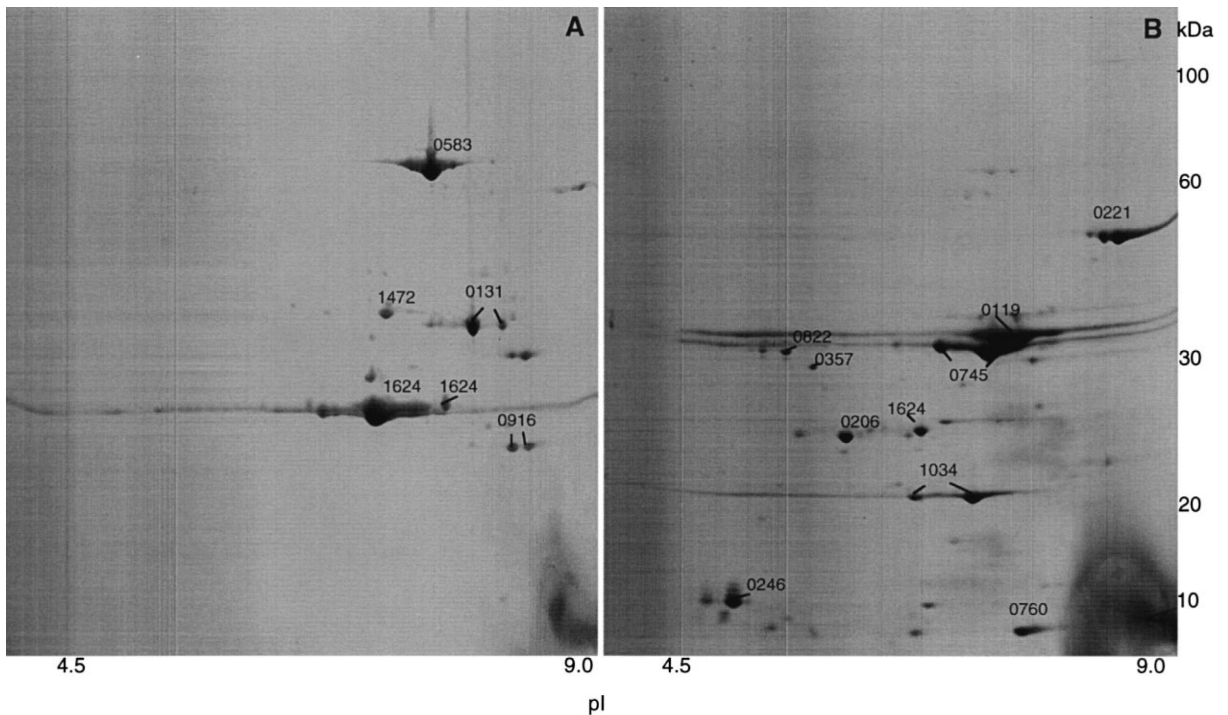

Fig. 3. 2D-PAGE analysis of the pools 11 (A) and 13 (B) of fractions eluted from the Polybuffer Exchanger with Polybuffer, pH 5.0. The column was developed and the proteins were analyzed as stated in the legend to Fig. 2. The horizontal stretching of strong spots (1624, 0221, 0119, 0745, 1034) is probably due to the presence of Polybuffer in the sample which affects the focusing of the proteins.

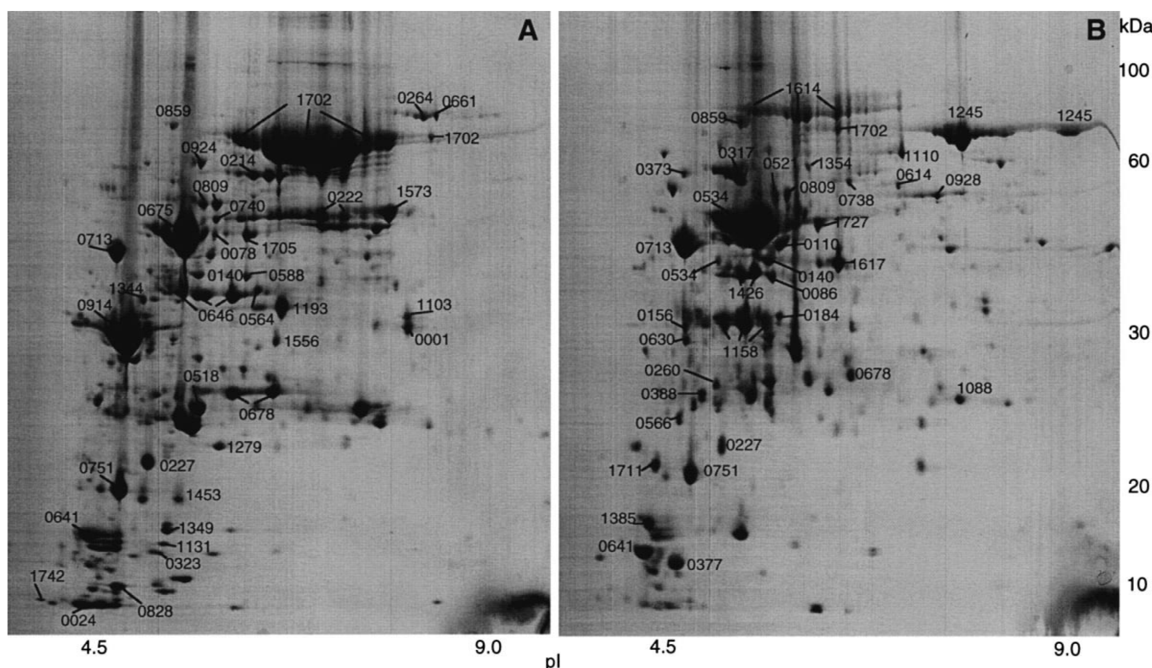

Fig. 4. 2D-PAGE analysis of the pools 17 (A) and 18 (B) of *H. influenzae* proteins eluted from the Polybuffer Exchanger with a NaCl gradient (0–2 *M*) following elution with Polybuffer, pH 5.0. The proteins were analyzed and the spots were numbered as stated in the legend to Fig. 2.

were strongly enriched. The recovery yields were not determined as we do not have specific assays for those proteins.

3.3. Protein enrichment by chromatofocusing

The pools that were eluted later with Polybuffer, included mixtures of proteins (Fig. 1B, pools 11–15). Several proteins were efficiently enriched and are represented by strong bands or spots on the 1D or 2D gels, respectively. The major enriched and partially purified proteins were: in pool 10, hypothetical protein HI0131, which was more than 70% pure (Fig. 1B, pool 10); in pool 11, hypothetical protein HI1624 (Fig. 1B, pool 11, the M_r 25 000, and Fig. 3A, the corresponding spot); in pool 12, D-galactose-binding periplasmic protein HI0822 (Fig. 1B, pool 12, the M_r 30 000 band); in pool 13, hypothetical protein HI0119 and arginine-binding periplasmic protein (HI0745) (Fig. 1B, pool 13, the bands migrating at about M_r 35 000, and Fig. 3B) and in pool 15, glyceraldehyde 3-phosphate dehydrogenase (HI0001) (Fig. 1B, pool 15, the M_r 30 000 band).

In the fractions eluted with the salt gradient more proteins were present. Pool 16 contained mainly phosphoenolpyruvate carboxykinase (HI0809), transaldolase (HI1125) and malonyl CoA acyl carrier protein (Fig. 1B, pool 16, the M_r 60 000, 32 000 and 25 000 bands, respectively). Some proteins were even stronger enriched, such as 5-methyltetrahydropteroyltriglutamate methyltransferase (HI1702) (Fig. 1B, pool 17, the M_r 70 000 band and Fig. 4A), peptidase D (HI0675), elongation factor Ts (HI0914) (Fig. 4A) and aspartase (HI0534) (Fig. 1B, pool 18, the M_r 50 000 band and Fig. 4B). Because of the significant enrichment, chromatofocusing can be used as an initial step for their isolation. In such a case, the column could be directly washed with Polybuffer, pH 5.0 and the above mentioned proteins can be subsequently eluted with a salt gradient.

The proteins bound to the Polybuffer Exchanger, which were eluted in the various pools (1–19), were identified from the 2D gels by MALDI-MS. The

identified proteins were located in a 2D map which shows the total soluble proteins of *H. influenzae* (Fig. 5). The proteins spots are designated with the *H. influenzae* (HI) identification numbers [1]. Their names are listed in Table 1. In previous articles a different labeling system was used [5–7]. However, in the protein lists of those articles, the HI numbers of the proteins identified were included, allowing thus an easy correlation with the data presented here. Table 1 is subdivided into four partial lists according to the degree of the protein enrichment by the chromatofocusing step in comparison with their abundance in the starting material. For some proteins found in the pools collected from the Polybuffer Exchanger, in particular acidic proteins, no clear correlation to a spot of Fig. 5 could be made. For these proteins the approximate locations on the 2D map are indicated.

4. Discussion

Visualization and identification of the proteins from microorganisms can facilitate the identification of new drug targets. However, not all proteins of a certain organism are expressed in amounts sufficient for detection by 2D-PAGE. In order to visualize and identify low-copy-number gene products of *H. influenzae*, we enriched proteins of the microorganism from larger volumes by applying various chromatographic techniques. Initial attempts with preparative electrofocusing were compromised by serious protein precipitation (data not shown). Chromatography on heparin–Actigel enriched many low-abundance proteins and in particular many basic ribosomal and other nucleic acid-binding proteins which were eluted with a high-salt buffer [5,18]. Here we investigated the protein enrichment by chromatofocusing. This step was chosen because the ion-exchanger has a high protein binding capacity and can discriminate and enrich proteins with minor differences in their isoelectric points (p*I* values).

The basic proteins with p*I* values higher than the pH 8.3 of the column equilibration buffer should theoretically not bind to the gel matrix and should be recovered in the flow-through and wash fractions. However, no significant amount of proteins were detected in them. In general, very few proteins with p*I* values at about 9, such as the major ferric iron-binding protein, were detected in the various pools. It is possible that most of the basic proteins were precipitated on the column during the sample application (a brown layer was formed with sample application which was not removed during column elution).

As expected, proteins with lower p*I* values were recovered during elution with Polybuffer. No clear separation of the proteins based solely on their p*I* values could be achieved. For example, in pool 13 proteins with p*I* values varying from approximately 5.0 to 9.0 were detected (Fig. 3B). It is possible that some of these proteins were associated to each other and eluted as a complex.

For the column development, we applied mild elution conditions, i.e., tenfold diluted Polybuffer, pH 5.0, in order to achieve an optimal resolution. Approximately 25% of the bound proteins were eluted from the column with this buffer [Table 1, the proteins designated with (P)]. Theoretically, all proteins with p*I* values higher than 5.0 should have been recovered under these conditions and this represents a much higher percentage of the *H. influenzae* proteins [8]. It seems that the buffering capacity of the diluted Polybuffer was not strong enough to displace strongly enriched proteins, such as 5-methyltetrahydropteroyltriglutamate methyltransferase (HI1702). These proteins as well as the proteins with p*I* values lower than approximately 5.5 were eluted with the salt-containing buffer, however, in a complex mixture with a number of other proteins (Fig. 4). The fractions eluted with salt contained the majority of the bound, mainly acidic proteins [Fig. 4 and Table 1, the proteins designated with (S)].

Comparison of Fig. 5 which shows the soluble proteins of *H. influenzae*, i.e., the starting material loaded onto the Polybuffer Exchanger, with the 2D gels of the pools collected from the column (which are partially shown here in Figs. 3 and 4) indicates that many of the enriched proteins were abundant proteins. Three out of the four very strongly enriched proteins, with the exception of peptidase D, (Table 1, list D) are represented by strong spots on Fig. 5. Approximately one fourth of the proteins bound were low-abundance gene products which became visible by chromatography on the Polybuffer exchanger and

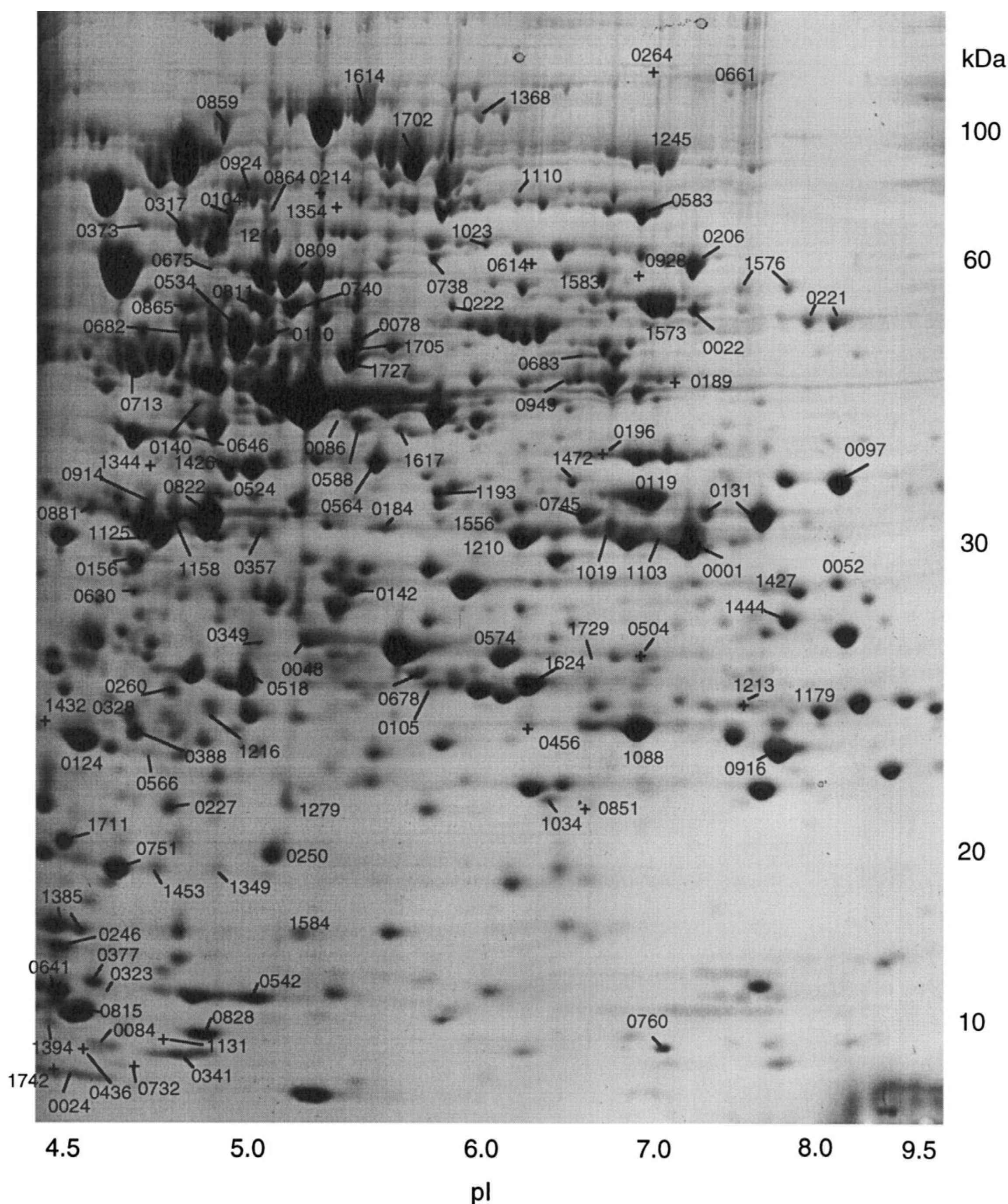

Fig. 5. Proteins of *H. influenzae* bound to Polybuffer Exchanger. The soluble proteins of *H. influenzae* were analyzed by 2D-PAGE as stated in Section 2 and in the legend to Fig. 2. The numbered spots represent proteins found in the pools collected from the column. For some of the proteins, which were enriched following chromatofocusing, no clear alignment to visible spots could be made, so that the approximate locations are indicated. The spots are labeled with the HI identification numbers. The names of the proteins are listed in Table 1. Proteins which were not found in the chromatofocusing fractions, although they may be known, are not indicated.

Table 1
Proteins of *H. influenzae* bound to Polybuffer Exchanger

HI number	Protein name	MALDI-MS		Location
		Peptides		(see Fig.)
		Matching	Total	
(A) Proteins bound to Polybuffer Exchanger but not significantly enriched				
HI0001	Glyceraldehyde 3-phosphate dehydrogenase (P, S)	7	23	4A, 5
HI0022	Citrate lyase α chain (P)	4	39	5
HI0048	Hypothetical protein HI0048 (S)	4	24	5
HI0097	Major ferric iron-binding protein (P)	5	15	2A, 5
HI0124	Probable inorganic pyrophosphatase (S)	7	9	5
HI0131	Hypothetical protein HI0131 (P)	11	19	3A, 5
HI0250	Single-strand-binding protein (S)	4	20	5
HI0317	Aspartyl-tRNA synthetase (S)	14	34	4B, 5
HI0328	Elongation factor P (S)	3	8	5
HI0357	Thiamin-repressed protein (P)	7	27	3B, 5
HI0373	Heat shock cognate (S)	7	27	4B, 5
HI0388	*H. influenzae* predicted coding region HI0388 (S)	4	15	4B, 5
HI0504	D-Ribose-binding periplasmic protein (P)	9	20	5
HI0524	Fructose-biphosphate aldolase (S)	7	17	5
HI0564	Asparagine synthetase A (S)	10	27	4A, 5
HI0574	Hypothetical protein HI0574 (P)	6	13	5
HI0579	Elongation factor G	12	28	5
HI0588	N-Carbamyl-L-amino acid amidohydrolase (S)	6	9	4A, 5
HI0641	50S ribosomal protein L7/L12 (S)	3	25	4A, 4B, 5
HI0682	Ketol-acid reductoisomerase (S)	10	23	5
HI0740	Hypothetical protein HI0740 (S)	17	33	4A, 5
HI0811	Argininosuccinate lyase (S)	6	17	5
HI0851	Hypothetical protein HI0851 (P)	3	22	5
HI0859	ATP-dependent protease binding protein (S)	9	21	4A, 4B, 5
HI0916	Export factor homolog (P)	5	11	3A, 5
HI0924	Glycyl-tRNA synthetase β chain (S)	13	26	4A, 5
HI1088	Superoxide dismutase (P)	5	21	4B, 5
HI1103	Cysteine synthase (S)	8	19	4A, 5
HI1179	Arginine-binding periplasmic protein (S)	4	6	5
HI1385	Ferritin-like protein (vsgA) (S)	3	17	4B, 5
HI1444	5,10 Methylene tetrahydrofolate reductase (P)	6	21	5
HI1472	*H. influenzae* predicted coding region HI1472 (P)	9	23	5
HI1556	Vancomycin resistance protein (S)	5	14	4A, 5
HI1573	Pyruvate kinase (S)	7	21	3A, 5
HI1576	Glucose-6-phosphate isomerase (P)	9	18	4A, 5
HI1583	Arginyl-tRNA synthetase (P)	14	21	5
HI1584	Acetolactate synthase III small (P)	5	21	5
HI1705	Aminopeptidase A/I (S)	11	17	4A, 5
HI1711	PTS system, glucose-specific IIA component (S)	5	17	4B, 5
(B) Proteins enriched by Polybuffer Exchanger chromatography				
HI0024	Citrate lyase γ chain (S)	3	20	5
HI0078	Cysteine-tRNA synthetase (S)	7	28	4A, 5
HI0084	Thioredoxin (P)	4	23	5
HI0105	Hypothetical protein HI0105 (S)	6	18	5
HI0119	Hypothetical protein HI0119 (P)	9	22	3B, 5
HI0140	N-Acetylglucosamine-6-phosphate (S)	13	29	4A, 4B, 5
HI0156	Malonyl CoA-acyl carrier protein (S)	6	15	4B, 5

(Cont.)

Table 1. Continued

HI number	Protein name	MALDI-MS		Location
		Peptides		(see Fig.)
		Matching	Total	
HI0184	Hypothetical protein HI0184 (S)	5	19	5
HI0189	Glutamate dehydrogenase (S)	8	24	5
HI0196	Chorismate synthase (P)	6	16	5
HI0221	Inosine-5′-monophosphate dehydrogenase (P)	12	26	3B, 5
HI0227	Hypothetical protein HI0227 (S)	5	24	4A, 4B, 5
HI0246	*H. influenzae* predicted coding region HI0246 (S)	4	22	3B, 5
HI0260	Hypothetical protein HI0260 (P)	4	17	4B, 5
HI0341	Hypothetical protein HI0341 (S)	2	18	4A, 5
HI0349	Adenylate kinase (S)	4	21	5
HI0436	Transformation protein (S)	3	14	5
HI0456	Hypothetical protein HI0456 (P)	9	24	5
HI0518	Purine nucleoside phosphorylase (S)	3	23	4A, 5
HI0583	2′,3′-Cyclic-nucleotide 2′-phosphodiesterase (P)	9	17	3A, 5
HI0614	L-Fucose isomerase (S)	5	19	4B, 5
HI0630	Mocoid status locus protein (S)	5	14	4B, 5
HI0661	Lactoferrin-binding protein (S)	7	23	4A, 5
HI0678	Triosephosphate isomerase (S)	5	19	4A, 4B, 5
HI0713	Trigger factor (S)	6	26	4A, 4B, 5
HI0732	*H. influenzae* predicted coding region HI0732 (S)	3	15	5
HI0738	Dihydroxyacid dehydratase (ilvD) (S)	7	20	4B, 5
HI0745	Probable L-asparaginase periplasmic protein (P)	11	21	3B, 5
HI0751	ToxR regulon (tagD) (S)	3	8	4A, 4B, 5
HI0815	Universal stress protein (P)	5	21	5
HI0864	Hypothetical M_r 68 400 GTP-binding protein (S)	4	17	5
HI0865	Glutamine synthetase (S)	8	22	5
HI0928	KW20 catalase (hktE) (S)	7	24	4B, 5
HI0949	γ-Aminobutyric acid transaminase (S)	8	33	5
HI1019	Thiamin-binding periplasmic protein (P)	6	19	5
HI1023	Transketolase I (S)	8	12	5
HI1110	D-Xylose transport ATP-binding protein (S)	3	13	4B, 5
HI1131	Cell division protein (ftsL) (S)	4	34	4A, 5
HI1210	Malate dehydrogenase (P)	4	7	5
HI1211	Lysyl-tRNA synthetase (S)	9	25	5
HI1213	Thiol:disulfide interchange protein (P)	6	13	5
HI1216	Hypothetical protein HI1216 (S)	8	25	5
HI1279	CMP-NeuNAc synthetase (S)	3	25	4A, 5
HI1348	Hypothetical protein HI1348 (S)	5	18	5
HI1349	Hypothetical protein HI1349 (P, S)	6	15	4A, 5
HI1354	Glutaminyl-tRNA synthetase (S)	6	10	4B, 5
HI1394	*H. influenzae* predicted coding region HI1394 (S)	3	34	5
HI1426	*H. influenzae* predicted coding region HI1426 (S)	9	17	4B, 5
HI1427	Hypothetical protein HI1427 (S)	8	19	5
HI1432	Tryptophan synthase α chain (S)	7	21	5
HI1453	Fimbrial transcription regulation repressor (S)	3	28	4A, 5
HI1614	Aminopeptidase N (S)	4	11	4B, 5
HI1742	RNA polymerase ω subunit (S)	3	15	4A, 5
HI1727	Argininosuccinate synthase (S)	6	18	4B, 5

Table 1. Continued

HI number	Protein name	MALDI-MS		Location
		Peptides		(see Fig.)
		Matching	Total	
(C) Proteins strongly enriched by Polybuffer Exchanger chromatography				
HI0052	Hypothetical protein HI0052 (P)	9	19	2B, 5
HI0086	Cystathionine γ synthase (S)	5	15	4B, 5
HI0104	Heat shock protein C62.5 (S)	11	27	5
HI0110	Seryl-tRNA synthetase (S)	7	16	4B, 5
HI0142	Probable N-acetylneuraminate lyase subunit (S)	5	14	5
HI0206	Probable 5′-nucleotidase (P)	10	20	2B, 2C, 5
HI0214	Oligopeptidase A (prlC) (S)	18	25	4A, 5
HI0222	GMP synthetase (S)	3	34	4A, 5
HI0264	Heme-hemopexin-binding protein (S)	15	26	4A, 5
HI0323	Hypothetical protein HI0323 (S)	3	20	4A, 5
HI0377	Nitrogen fixation protein (S)	5	20	4B, 5
HI0498	Spermidine/putrescine-binding protein (S)	7	21	5
HI0542	M_r 10 000 Chaperonin (GroES) (S)	3	25	5
HI0566	Dod protein (S)	4	22	4B, 5
HI0646	Aspartate-semialdehyde dehydrogenase (S)	7	15	4A, 5
HI0760	Unknown protein HI0760 (P)	3	24	3B, 5
HI0809	Phosphoenolpyruvate carboxykinase (S)	16	24	4A, 4B, 5
HI0822	D-Galactose-binding periplasmic protein (P)	10	22	3B, 5
HI0828	Potassium channel homolog (S)	3	22	4A, 5
HI0881	Hypothetical protein HI0881 (S)	6	17	5
HI1034	Hypothetical protein HI1034 (P)	6	20	3B, 5
HI1125	Transaldolase (S)	7	20	5
HI1158	Thioredoxin reductase (S)	4	16	4B, 5
HI1193	Branched-chain amino acid transaminase (S)	6	20	4A, 5
HI1245	Putative malate dehydrogenase (S)	4	5	4B, 5
HI1344	Spermidine/putrescine-binding protein (S)	10	27	4A, 5
HI1368	Hypothetical protein HI1368 (S)	3	9	5
HI1617	Aspartate aminotransferase (S)	5	16	4B, 5
HI1624	Hypothetical protein HI1624 (P)	5	18	3A, 5
(D) Proteins very strongly enriched by Polybuffer Exchanger chromatography				
HI0534	Aspartate ammonia lyase (S)	11	22	4B, 5
HI0675	Peptidase D (S)	4	11	4A, 5
HI0914	Elongation factor ET-Ts (S)	5	16	4A, 5
HI1702	5-Methyltetrahydropteroyltriglutamate methyltransferase (S)	10	19	4A, 4B, 5

The column was eluted with Polybuffer, pH 5.0 and afterwards with a NaCl gradient. The fractions were pooled according to the elution profile (Fig. 1A) and the pools were analyzed by 2D-PAGE (Figs. 2–4). The protein spots from the 2D gels were analyzed by MALDI-MS. The number of total and matching peptides derived from MS are given. In the column ''Location'', the Figure number is indicated in which the corresponding protein spot can be found. (P) and (S) next to the protein name indicate whether the protein was eluted with Polybuffer, pH 5.0 (pools 1–15) or salt (pools 16–19), respectively. The spots or approximate locations of all bound proteins are shown in Fig. 5. Proteins which were not found in the chromatofocusing fractions are not listed in this Table and are not shown in Fig. 5.
HI, *H. influenzae* identification number, according to Fleischmann et al. [1].

could be identified and approximately localized on the 2D map (Fig. 5).

About 125 proteins were identified in the pools collected from the Polybuffer Exchanger (Table 1). Additional proteins in the pools, represented mainly by minor spots, were not identified and are not included in Table 1. In order to search for more basic proteins, the pools were also analyzed by immobilized pH gradient strips comprising the region 6–11. No additional proteins were found on these gels (data

not shown). Table 1 includes about 70 newly identified proteins. Many of them are not represented by a visible spot on the 2D map comprising the soluble proteins, so that they could not have been identified had they not been enriched by chromatofocusing (the proteins with weak or nonvisible signals in Fig. 5). Together with the identified proteins described here, the total number of identified and published proteins of *H. influenzae* increases to about 360 (Refs. [5–7,18] and this article). Chromatofocusing is not a reliable method to determine the recovery of the proteins enriched by this step, as we do not know the expression levels in the starting material and the chromatographic step is associated with side-events such as precipitation during chromatography. Protein quantification by other approaches is in progress.

The proteins bound to Polybuffer Exchanger can not be clustered into one class. Most of them are enzymes with various functions (Table 1). Table 1 includes many proteins which in the database are designated as hypothetical proteins or as predicted coding regions (unknown proteins). The detection of the hypothetical proteins in the proteome of *H. influenzae* provides the evidence that these proteins are really expressed, some of them in high copy numbers, as for example the proteins HI0119 and HI1624 (Fig. 3). The potential functions of these proteins can be deduced from homology to known counterparts from other microorganisms by search in databases.

Many proteins are represented by more than one spots, such as the proteins HI0052 (Fig. 2B), HI0131 (Fig. 3A), HI1034 (Fig. 3B), HI1702 (Fig. 4A), HI1614 (Fig. 4B) and others. The reasons of the heterogeneity may be protein carbamylation taking place during the performance of the 2D gel electrophoresis, posttranslational modifications, such as acylation, phosphorylation or N-terminal modifications, or focusing artifacts when large amounts of one major component are present (for example protein HI1702, Fig. 4A). The protein heterogeneity on 2D gels is currently under investigation.

Making use of recent improvements in the daily sample analysis throughput by MS [13] and sequence information of the genomes of various microorganisms, we were able to identify most proteins present in each pool of the Polybuffer Exchanger. We provide here a systematic catalog of the proteins of *H. influenzae* collected at the various stages of a chromatographic process. The list of proteins bound to Polybuffer Exchanger may be a useful guide to Biochemists aiming to isolate such proteins from the same or other bacterial microorganisms. Bacteria, such as *Escherichia coli* [19], show a distribution profile of their proteins on a 2D map similar to that of *H. influenzae* (Fig. 5). Similar purification schemes can possibly be applied for the isolation or enrichment of homologous proteins.

Two lists with the proteins of *H. influenzae* detected in the eluates from the heparin–Actigel [5] and Polybuffer Exchanger are now available. One can combine the two fractionation approaches and possibly achieve a high purity for many proteins. For example, elongation factor G (HI0579) was recovered in the pool of fractions 27–30 eluted from the heparin gel [5] and in the pool 19 from the Polybuffer Exchanger, whereas practically none of the other proteins present in one was present in the other pool (pool 19 not shown). Similarly, 5-methyltetrahydropteroyltriglutamate methyltransferase (HI1702) was present in fractions 21–24 of the heparin column [5] and in pool 17 from the Polybuffer Exchanger (Fig. 4A). Sequential chromatography on the two gel matrix types should remove most of the other proteins [trigger factor (HI0713) was present in both pools]. Similar purification examples can be given for other proteins as well, such as phosphoenolpyruvate carboxykinase (HI0809), D-galactose-binding periplasmic protein (HI0822) etc. (the 2D gel images with the proteins present in each Polybuffer Exchanger pool are available on request).

In addition to the easier design of purification protocols, the use of selected chromatographic steps before the 2D gels analysis can significantly facilitate the analysis of complex protein mixtures. The chromatography methods separate complex protein mixtures into simpler fractions on the basis of different binding principles and every approach adds a unique resolving power which can be essential for an efficient proteome analysis. A combination of sequential chromatographic steps can be even more powerful. A sequence of analytical approaches, comprising multiple chromatographic steps, 2D-PAGE analysis and MS, is indispensable for the proteome mapping of higher organisms.

In summary, we fractionated the total soluble proteins of the bacterium *H. influenzae* by chromatography on Polybuffer Exchanger. The protein pools were analyzed by 2D gels and the proteins identified by MALDI-MS. Two proteins were purified to homogeneity, another one was obtained in a pure form and four others were partially purified. Seventy new proteins were identified in the chromatofocusing pools of which about half were low-abundance gene products. The results may be useful in the search for drug targets and in the design of purification protocols for isolation of proteins of interest from *H. influenzae* and other microorganisms.

Acknowledgements

We thank J.-F. Juranville and M.-F. Takács for technical assistance and Drs. H. Lötscher and J. Mous for their support.

References

[1] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.-F. Tomb, B.A. Dougherty, J.M. Merrick, K. Kenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.I. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, J.C. Venter, Science 269 (1995) 496.
[2] L. Anderson, Science 270 (1995) 369.
[3] P. Cash, J. Chromatogr. A 698 (1995) 203.
[4] D.S. Young, R.P. Tracy, J. Chromatogr. A 698 (1995) 163.
[5] M. Fountoulakis, H. Langen, S. Evers, C. Gray, B. Takacs, Electrophoresis 18 (1997) 1193.
[6] M. Fountoulakis, J.-F. Juranville, P. Berndt, Electrophoresis 18 (1998) 2968.
[7] H. Langen, C. Gray, D. Roeder, J.-F. Juranville, B. Takacs, M. Fountoulakis, Electrophoresis 18 (1997) 1184.
[8] H. Langen, D. Roeder, J.-F. Juranville, M. Fountoulakis, Electrophoresis 18 (1997) 2085.
[9] L.A.Æ. Sluyterman, O. Elgersma, J. Chromatogr. 150 (1978) 17.
[10] L. Giri, Methods Enzymol. 182 (1990) 380.
[11] M. Fountoulakis, E.J. Schlaeger, R. Gentz, J.-F. Juranville, M. Manneberg, L. Ozmen, G. Garotta, Eur. J. Biochem. 198 (1991) 441.
[12] B. Bjellqvist, C. Pasquali, F. Ravier, J.-C. Sanchez, D.F. Hochstrasser, Electrophoresis 14 (1993) 1357.
[13] M. Fountoulakis, H. Langen, Anal. Biochem. 250 (1997) 153.
[14] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.
[15] C.H. Wheeler, S.L. Berry, M.R. Wilkins, J.M. Corbett, K. Ou, A.A. Gooley, I. Humphery-Smith, K.L. Williams, M.J. Dunn, Electrophoresis 17 (1996) 580.
[16] M.R. Wilkins, C. Pasquali, R.D. Appel, K. Ou, O. Golaz, J.-C. Sanchez, J.X. Yan, A.A. Gooley, G. Hughes, I. Humphery-Smith, K.L. Williams, D.F. Hochstrasser, BioTechnology 14 (1996) 61.
[17] P.G. Righetti, A. Bossi, Anal. Biochem. 247 (1997) 1.
[18] M. Fountoulakis, B. Takacs, H. Langen, Electrophoresis (1998) in press.
[19] C. Pasquali, S. Frutiger, M.R. Wilkins, G.J. Hughes, R.D. Appel, A. Bairoch, D. Schaller, J.C. Sanchez, D.F. Hochstrasser, Electrophoresis 17 (1996) 547.